# From Safety-I to Safety-II: A brief introduction to resilience engineering

Erik Hollnagel
Professor, University of Southern Denmark
Chief Consultant, Centre for Quality, Region of Southern Denmark
hollnagel.erik@gmail.com

*Perceive those things which cannot be seen*
*Miyamoto Musashi (1584 – 1645)*

## Introduction

Safety is traditionally defined as a condition where nothing goes wrong. Or rather, since we know that it is impossible to ensure that nothing goes wrong, as a condition where the number of things that go wrong is acceptably small. This can be seen from the following definitions:

- Safety is the freedom from unacceptable risk. (The American National Standards Institute).
- Safety is the freedom from accidental injury. (U.S. Agency for Healthcare Research and Quality).
- Safety is the state in which harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and risk management. (International Civil Aviation Organization).

In relation to human activity it makes good practical sense to focus on situations where things go wrong, both because such situations by definition are unexpected and because they may lead to unintended and unwanted harm or injury. Such events were initially explained as 'acts of god' or 'acts of nature', and therefore as beyond human control. (Ironically, the 'acts of nature' seem to be returning with a vengeance.) This view changed as humans gradually became masters of technology, especially after the second industrial revolution around 1750. The rapid mechanisation of work that followed had, as a side effect, a growing number of accidents, where the common factor was the breakdown, failure, or malfunctioning of technology. Safety concerns therefore focused on guarding machinery, stopping explosions and preventing structures from collapsing. The focus on technology as the main source of both problems and solutions was successfully maintained until 1979, when the accident at the Three Mile Island nuclear power plant demonstrated that safeguarding technology was not enough. The TMI accident put the spotlight on the human factor and made it necessary to consider human failure and malfunctioning as a potential risk. In 1986, only seven years later, the loss of the space shuttle Challenger, reinforced by the accident in Chernobyl, required yet another change, this time to include the influence of organisational failures and safety culture to the common lore.

Historically speaking, new types of accidents have been matched by new types of causes (e.g., metal fatigue, 'human error,' organisational failure, complex systems) without challenging the underlying assumption of causality. We have therefore become so used to

explain accidents in terms of cause-effect relations that we no longer notice it. And we cling tenaciously to this tradition, although it has becomes increasingly difficult to reconcile with reality.

## The Causality Credo

Any explanation of how accidents happen must include some assumption about how causes lead to effects. This is often called an accident model. The simplest accident model is the Domino model (Heinrich, 1931), although this way of thinking probably is as old as mankind itself. The Domino model represents simple linear causality using the analogy of a set of domino pieces that fall one after the other. According to the logic of these models, the purpose of event analysis is to reason backwards from the injury to find the 'root cause'. Similarly, risk analysis looks for whether something may 'break', meaning that a specific component may fail or malfunction, either by itself or in combination with another failure or malfunction.

Simple, linear models were superseded in the 1980s by the composite linear models, where the best known example is the Swiss cheese model. According to these models adverse outcomes can be explained as combinations of active failures (or unsafe acts) and latent conditions (hazards). An event analysis thus looks for how degraded barriers or defences can combine with active (human) failures. Similarly, risk analysis focuses on finding the conditions under which combinations of single failures and latent conditions may result in an adverse outcome, where the latent conditions are conceived of as degraded barriers or weakened defences.

The Domino model and the Swiss cheese model are typical examples of accident models, but many others exist. Common to them all is the unspoken assumption that outcomes can be understood as effects that follow from prior causes. Since that corresponds to a belief – or even a faith – in the laws of causality, it may be called a causality credo. The causality credo can be expressed as follows:

- Adverse outcomes happen because something has gone wrong. Adverse outcomes have causes.
- It is possible to find these causes provided enough evidence is collected. Once the causes have been found, they can be eliminated, encapsulated, or otherwise neutralised.
- Since all adverse outcomes have causes, and since all causes can be found, it follows that all accidents can be prevented. This is the vision of zero accidents or zero harm that many companies find attractive.

While reasoning in this manner may be plausible for systems that are relatively simple, it does not suffice for more complicated systems. And since most systems today are complicated rather than uncomplicated, as recognised by Perrow already in 1984, the causality credo is no longer sustainable.

## Safety-I: Avoiding That Things Go Wrong

The historical development of safety thinking in combination with the causality credo leads to the view of safety illustrated by Figure 1. According to this, unacceptable outcomes happen because of preceding failures and malfunctions, while acceptable outcomes happen because everything – including people – worked as it should. This corresponds to a

'*hypothesis of different causes*,' which states that the causes or 'mechanisms' of adverse events are different from those of events that succeed. If that was not the case, the elimination of such causes and the neutralisation of such 'mechanisms' would also reduce the likelihood that things could go right, hence be counterproductive.

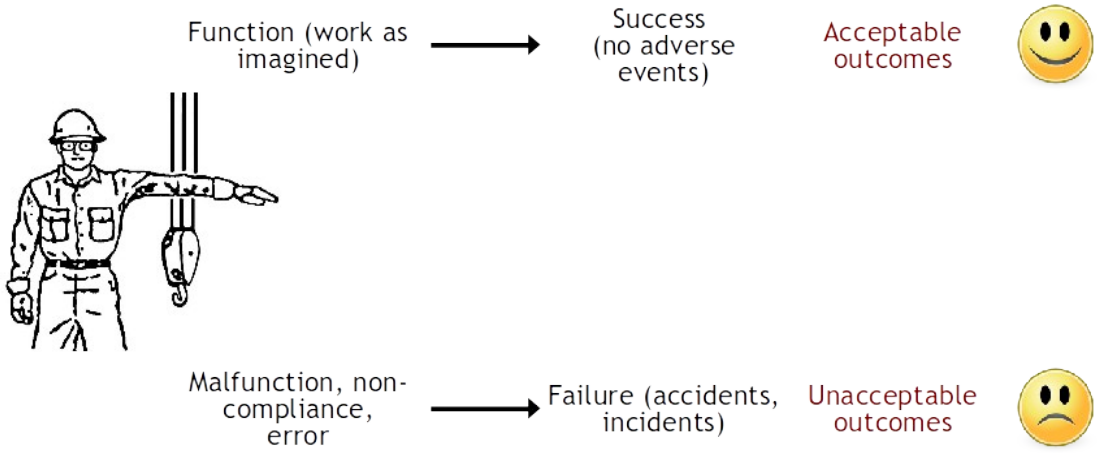


*Figure 1: The Safety-I view of failures and successes*

Safety is traditionally defined as a condition where the number of adverse outcomes (accidents / incidents / near misses) is as low as possible. This can be called Safety-I. The purpose of safety management is consequently to achieve and maintain that state. While this definition is simple, it is problematic because safety is defined by its opposite, by what happens when it is missing. It is also measured indirectly, not by its presence or as a quality in itself, but by the consequences of its absence.

Safety-I tacitly assumes that systems work because they are well designed and scrupulously maintained, because procedures are complete and correct, because designers can foresee and anticipate even minor contingencies, and because people behave as they are expected to – and more importantly as they have been taught or trained to do. This unavoidably leads to an emphasis on *compliance* in the way work is carried out.

## Looking at what goes wrong rather than looking at what goes right

Resilience engineering argues that the Safety-I perspective is both oversimplified and wrong. Resilience engineering rejects the hypothesis of different causes and instead propose that things that go right and things that go wrong happen in basically the same way (Hollnagel et al., 2006 & 2012). This means that we cannot understand how unacceptable outcomes happen unless we first understand how acceptable outcomes happen. To illustrate the consequences of looking at what goes wrong rather than looking at what goes right, consider Figure 2. This represents the case where the (statistical) probability of a failure is 1 out of 10,000. In other words, for every time we expect that something will go wrong (the red line), there are 9,999 times where we should expect that things will go right and lead to the outcome we want (the green area).

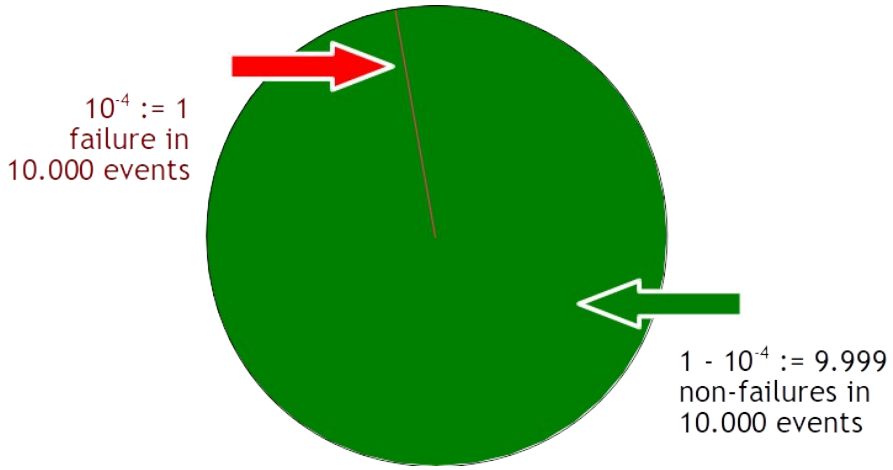


*Figure 2: The imbalance between things that go right and things that go wrong*

The focus on what goes wrong is required by regulators and authorities, supported by models and methods, documented in countless databases, and described in literally thousands of papers, books, and conference proceedings. The net result is a deluge of information both about how things go wrong and about what must be done to prevent this from happening. The recipe is the simple principle known as 'find and fix': look for failures and malfunctions, try to find their causes, and try to eliminate causes and/or improve barriers.

The situation is quite different when it comes to that which goes right, i.e., the 9,999 events out of the 10,000. The focus on what goes right receives little encouragement; it is not required by authorities; there are few theories or models about how human and organisational performance succeeds, and few methods to help us study how it happens; actual data are difficult to locate; it is hard to find papers, books or other forms of scientific literature about it; and there are few people who even consider it worthwhile. In other words, we spend a lot of effort to understand why things go wrong, but very little effort to understand why they go right. We study the absence of safety rather than the presence of safety!

## Why do things go right?

Resilience engineering proposes that systems work because people are able to adjust what they do to match the conditions of work. People learn to identify and overcome design flaws and functional glitches, they can recognise the actual demands and adjust their performance accordingly, and they interpret and apply procedures to match the conditions. People can also detect and correct when something is about to go wrong, hence intervene before the situation becomes grave. This can be described as performance variability, not in the negative sense of deviations from some norm or standard, but in the sense of the smooth adjustments that are necessary for safety and productivity.

Performance variability or performance adjustments are a *sine qua non* for the functioning of today's socio-technical systems. Unacceptable outcomes or failures can therefore not be prevented by eliminating or constraining performance variability since that would also affect the desired acceptable outcomes. Instead efforts are needed to facilitate the necessary performance adjustments by clearly representing resources and constraints of a situation

and by making it easier to anticipate the consequences of actions. Performance variability should be managed by attenuation (dampening) if it seems to go in the wrong direction and by strengthening (amplification) if is seems to go in the right direction. In order to do so it is necessary first to acknowledge the inevitability and necessity of performance variability, second to monitor it, and third to control it.

## Safety-II: Ensuring That Things Go Right

Our socio-technical systems continue to develop and become more complicated, not least due to the allure of ever more powerful information technology. The models and methods of Safety-I are therefore increasingly unable to deliver the required and coveted 'state of safety'. Rather than 'stretching' the tools of Safety-I even further, we can change the definition of safety from 'avoiding that something goes wrong' to 'ensuring that everything goes right' – or more precisely to the ability to succeed under varying conditions, so that the number of intended and acceptable outcomes (in other words, everyday activities) is as high as possible. This can be called Safety-II (Figure 3). The basis for safety and safety management now becomes an understanding why things go right, which means an understanding of everyday activities.

Because everything basically happens in the same way regardless of the outcome, it is no longer necessary need to have different causes and 'mechanisms' for things that go wrong (accident and incidents) and for things that go right (everyday work). The purpose of safety management is to ensure latter, and by doing so it will also reduce the former. Safety-I and Safety-II therefore both lead to a reduction in unwanted outcomes, but use fundamentally different approaches with important consequences for how the process is managed and measured – as well as for productivity and quality.

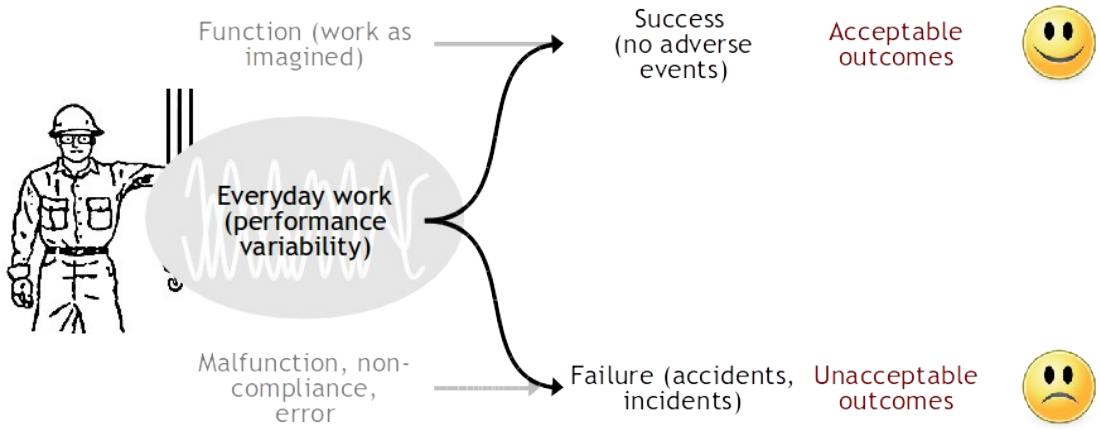


*Figure 3: The Safety-II view of failures and successes*

From a Safety-II perspective, the purpose of safety management is to ensure that as much as possible goes right and that everyday work achieves its stated purposes. This cannot be done by responding alone, since that will only correct what has happened. Safety management must instead be proactive. For this to work, it is necessary to foresee what could happen with acceptable certainty and to have the appropriate means (people and resources) to do something about it. That in turn requires an understanding of how the

system works, of how its environment develops and changes, and of how functions may depend on and affect each other. This understanding can be developed by looking for patterns and relations across events rather than for causes of individual events. To see and find those patterns, it is necessary to take time to understand what happens rather than spend all resources on fire-fighting.

## Conclusion

By juxtaposing Safety-I and Safety-II it becomes clear what the consequences are of basing safety management on one or the other (Table 1).

| | Safety-I | Safety-II |
|---|---|---|
| Definition of safety | That as few things as possible go wrong | That as many things as possible go right |
| Safety management principle | Reactive, respond when something happens | Proactive, try to anticipate developments and events |
| Explanations of accidents | Accidents are caused by failures and malfunctions | Things basically happen in the same way, regardless of the outcome. |
| View of the human factor | Liability | Resource |

While the development from a Safety-I approach to a Safety-II approach will neither be simple or fast, some practical suggestions for how to begin are given below:

- *Look at what goes right, as well as what goes wrong*. Things go well because people make sensible adjustments according to the demands of the situation. Find out what these adjustments are and try to learn from them!
- *When something has gone wrong, look for everyday performance variability rather than for specific causes*. Whenever something is done, it is a safe bet that it has been tried before. People quickly learn which performance adjustments work and soon come to rely on them – precisely because they work. Blaming people for doing what they usually do is therefore counterproductive.
- *Look at what happens regularly and focus on events based on how often they happen (frequency) rather than how serious they are (severity)*. It is much easier to be proactive for that which happens frequently than for that which happens rarely. A small improvement of everyday performance may count more than a large improvement of exceptional performance.
- *Allow time to reflect, to learn, and to communicate*. If all the time is used trying to make ends meet, there will no time to consolidate experiences or replenish resources – including how the situation is understood.
- *Remain sensible to the possibility of failure – and be mindful*. Try to think of undesirable situations and imagine how they may occur. Then think of ways in which they can either be prevented from happening, or be recognised and responded to as they are happening. This is the essence of proactive safety management.

Since the socio-technical systems on which our existence depends continue to become more and more complicated, remaining with a Safety-I approach will be inadequate in the long run. Yet the way ahead does not lie in a wholesale replacement of Safety-I by Safety-

II, but rather in a combination of the two ways of thinking. Safety-II is first and foremost a different understanding of what safety is, hence also a different way of applying many of the familiar methods and techniques. In addition to that it will also require methods on its own, to look at things that go right, to analyse how things work, and to *manage* performance variability rather than just *constraining* it (Hollnagel, 2013). We cannot make things go right simply by preventing them from going wrong. We can only make things go right by understanding the nature of everyday performance and by learning how to perceive those things which we otherwise do not see.

## References

Heinrich, H. W. (1931). Industrial accident prevention: A scientific approach. McGraw-Hill.

Perrow, C. (1984). *Normal Accidents.* New York: Basic Books.

Hollnagel, E. (2013). FRAM: The functional resonance analysis method for modelling complex socio-technical systems. Farnham, UK: Ashgate.

Hollnagel, E., Woods, D. D. & Leveson, N. (Eds.) (2006). Resilience engineering: Concepts and precepts. Farnham, UK: Ashgate.

Hollnagel, E., Paries, J., Woods, D. D. & Wreathall, J. (Eds.) (2011). Resilience engineering in practice: A guidebook. Farnham, UK: Ashgate.